INTEGRATED BIOINFORMATICS ANALYSIS OF PROTEOMICS IN PLANTS IN RESPONSE TO SALT STRESS

Submitted by

MOULI NAHAR, M.Sc.

Registration Number: SVU/PhD/SCI/BT/006/2021; 14.05.2022

Under the supervision of

DR. SEMANTI GHOSH

ASSISTANT PROFESSOR

DEPARTMENT OF BIOTECHNOLOGY

SWAMI VIVEKANANDA UNIVERSITY

THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN SCIENCE

(BIOTECHNOLOGY)

OF

SWAMI VIVEKANANDA UNIVERSITY

2025

Abstract

Asian rice (*Oryza sativa* L.) is a primary staple food worldwide; however, it faces significant challenges from salinity stress. This detrimental abiotic factor not only impedes growth but also markedly reduces yield and productivity. Despite efforts by researchers to enhance salt tolerance in rice through the identification of relevant biomarkers, the underlying mechanisms of this tolerance remain unclear, largely due to the involvement of multiple associated genes. In this study, we aimed to investigate and identify the critical genes associated with salinity stress in *Oryza* using computational bioinformatic analysis of gene expression. A diverse no. of differential expression genes (DEGs) have been identified from GSE58603, P 0.05 and |log FC > 1.0. Gene ontology (GO) and pathway enrichment analysis were performed. GO theories are divided into molecular function (MF), cellular components (CC), signalling pathways, and biological processes (BP). Ten hub genes were identified, suggesting their potential involvement in the response to salt stress. The analysis of upregulated genes and their proteinprotein interactions highlighted significant hub genes that are fundamentally linked to the onset of salinity stress. The microarray gene analysis revealed one of the most noted upregulated and down regulated genes respectively as allene oxide synthase and dehydrin Rab16. According to various studies, PLAT and Dehydrin are the two major proteins known to combat salinity stress as well as other associated stresses in rice. While dehydrin being relatively well studied previously, PLAT was not. This study revealed sequence similarities in PLAT-1 between *Arabidopsis* and *Oryza* spp., indicating a possible shared mechanism in response to salinity stress. The developed models may provide a foundation for future research in this area. During the homology model building, PLAT-1 in both Arabidopsis and Oryza exhibited relatively high shared similarities with the allene oxide synthase-lipoxygenase (AOS) protein. Previous studies have shown that AOS acts actively towards JA biosynthesis and biotic and abiotic stresses, typically by forming the LOX2AOS-AOC2 complex. This observation raises the question of whether AOS and PLAT can function synergistically in plant stress defence. MYB6 from *O. sativa* has been analysed also. The MYB TF is a highly conserved class of TF and plays a role in stress tolerance in rice. To evaluate the OsMYB6 protein in rice and other plant species, a phylogenetic tree was constructed. Conserved protein domains and motif analysis results predicted the fundamental role of TF functionality. However, gene structure analysis revealed variations among species that may predict different downstream

regulatory genes that participate in multiple signalling pathways related to salt stress. We targeted the major genes and protein interactors in miR166 are related to stress, more specifically, salinity stress tolerance in *Oryza sativa*. According to numerous previous studies, miR166 typically targets the regulatory gene MYB6 to silence phenotypic advancement in plants during stress. This advancement could only be gained if this silencing phenomenon be avoided. AGO1 protein well known for its silencing mechanism for miRNA. In this study, through molecular docking method the gene silencing of *mir166* through AGO1 protein has been done. Gene silencing of miRNA can be a new way to mask its effect, while letting flourish the expression of important proteins/genes that involved in salt stress tolerance in rice.

Consequently, improving the crop yield quality is of significant importance.

Chapter 1: Identification of Key Genes Related to Salinity Stress in Rice (*Oryza sativa*) Using Bioinformatics Analysis

Methodology:

Retrieval of Gene Expression Datasets

This study focused on elucidating the molecular response of *Oryza sativa* (rice) to salinity stress through comprehensive transcriptomic analysis. Gene expression datasets were retrieved from the Gene Expression Omnibus (GEO) database, with the series GSE58603 selected for its detailed profiling of salinity-stressed rice samples. This dataset provided high-quality transcriptomic data suitable for downstream analyses.

Data Analysis of DEGs

Differential gene expression analysis was conducted using GEO2R, comparing three biological replicates of salt-stressed seedlings against three controls. Genes with a P-value < 0.05 and absolute log fold change ≥ 1 were classified as differentially expressed (DEGs), distinguishing upregulated and downregulated genes responsive to salt stress.

Gene Functional and Pathway Enrichment Analysis

Functional annotation and pathway enrichment analyses were performed using ShinyGO, leveraging Gene Ontology (GO) and KEGG databases to identify significantly enriched

biological processes, molecular functions, cellular components, and pathways. This step revealed key molecular mechanisms and cellular pathways implicated in the salt stress response.

Building A PPI Network and Detection of Key Genes

To explore protein-level interactions, DEGs were mapped onto the STRING database to construct a high-confidence protein-protein interaction (PPI) network (combined score > 0.7).

The network was visualized and analysed in Cytoscape, where centrality metrics were calculated using plugins such as cytoHubba and Network Analyzer. These analyses identified hub genes with critical roles in network stability and information flow, highlighting potential key regulators of the salinity stress response in rice.

Result

Salt stress in rice roots primarily upregulates genes related to stress response, transport, and photosynthesis, with strong enrichment in transporter and oxidoreductase functions. Salt & ABA combined stress shifts the response, resulting in fewer upregulated but more downregulated key genes, with a focus on redox processes and nitrogen metabolism. Hub genes identified via PPI networks suggest central roles for CCR1 (salt) and NIA1 (salt & ABA) in the stress response.

Leaf tissue shows strong upregulation of classic stress-responsive genes under salt stress. Salt and salt & ABA treatments trigger distinct gene expression and functional enrichment profiles in rice seedlings, highlighting different molecular strategies for coping with abiotic stress. The addition of ABA notably alters the pattern of gene regulation, especially increasing the downregulation of previously upregulated stress-responsive genes.

Chapter 2: *In silico* analysis: upregulated PLAT domain-containing protein 1 during salt stress in *Oryza sativa*

Methodology:

Sequence Extraction and Alignment

PLAT domain-containing plant proteins linked to abiotic stress responses, especially salinity and temperature stress, were identified through comprehensive literature review. Experimentally validated protein sequences were retrieved in FASTA format from the UniProtKB/Swiss-Prot database to ensure data reliability. A custom BLAST database was created from these sequences to facilitate homology searches, enabling identification of conserved regions and functional motifs across various plant species. Multiple sequence alignment was performed using CLUSTAL-OMEGA with default settings to detect conserved domains and sequence variations relevant to stress tolerance. The alignment results were analysed to correlate conserved motifs with known functional roles of PLAT proteins in abiotic stress tolerance

Protein 3D Modelling

Since the 3D structure of PLAT-1 protein was unavailable in the Protein Data Bank, homology models for Arabidopsis thaliana and Oryza sativa PLAT-1 proteins were constructed. PHYRE2 was initially attempted but was unsuccessful due to low sequence identity, leading to the use of comparative modelling via the SWISS-MODEL server. The generated models were validated using the SAVES server to assess stereochemical quality and residue geometry. Ramachandran plot analysis was conducted using Ramplot for 2D and 3D structural mapping. Structural superimposition between homology models and template structures was performed using the SUPERPOSE server to evaluate model accuracy and structural similarity.

Protein-Protein Interaction

Potential protein-protein interactions involving PLAT-1 proteins from Oryza species were explored using the STRING database, which integrates experimental and predicted interaction data. Protein sequences or gene IDs were submitted specifying Oryza as the organism to obtain species-specific interaction networks. A high-confidence threshold (0.900) was applied to ensure reliability of the predicted interactions. The resulting networks were analysed to identify functional partners and to infer biological roles of PLAT-1 proteins in abiotic stress responses in rice. This methodology integrates sequence analysis, structural modelling, and interaction network prediction to comprehensively characterize PLAT domain proteins involved in plant

abiotic stress tolerance, providing insights into their evolutionary conservation, structural features, and functional associations.

Result

The PLAT-1 protein is highly conserved between *Arabidopsis thaliana* and *Oryza sativa*, especially in its central region, suggesting similar functions. Reliable 3D models were built and validated for both species, providing a structural basis for future research into their roles in plant biology.

Chapter 3: In-silico analysis of *OsMYB6* transcription factor in *Oryza sativa* conferring salinity stress

Methodology:

Sequence Retrieval

Genomic, coding (CDS), and protein sequences for twelve dicot and ten monocot species were retrieved from NCBI, Ensembl Plants, and UniProt databases. Chromosomal positions and strand orientations were also obtained to facilitate downstream analyses.

Protein Domain and Motif Analysis

Conserved MYB-DNA binding domains were identified using SMART and NCBI CD-search tools, with a focus on PFAM domain detection. The MEME Suite was utilized to discover up to ten statistically significant conserved motifs within the MYB protein sequences, revealing sequence patterns likely linked to functional conservation.

Phylogenetic and Multiple Sequence Alignment

Evolutionary relationships among OsMYB and homologous genes were inferred using ClustalW for sequence alignment and the Neighbor-Joining method in MEGA11 for phylogenetic tree construction. Robustness was assessed via 1000 bootstrap replicates.

Multiple sequence alignments were refined and annotated with GeneDoc to highlight conserved regions.

Gene Structure Analysis

The exon-intron organization of MYB6 genes was visualized using the GSDS 2.0 server by aligning CDS with genomic DNA. Integration of the phylogenetic tree enabled comparative visualization of gene architectures, with exons, introns, and regulatory regions distinctly represented.

Protein Physicochemical Properties

ProtParam (ExPASy) was used to estimate protein length, molecular weight, isoelectric point, and amino acid composition, providing a detailed physicochemical profile for each MYB protein.

Subcellular Localization

DeepLoc 2.0, a deep learning-based predictor, was employed to determine the probable subcellular localization of MYB proteins, supplemented by UniProt annotations for validation.

Result

The study comprehensively characterized the MYB6 gene and protein from rice and compared it with homologs from other plant species. The results demonstrate high conservation of key domains and motifs, evolutionary relationships, structural similarities, and differences, as well as diverse physio-chemical properties and subcellular localizations. These findings provide valuable insights into the structure, function, and evolution of MYB transcription factors in plants.

Chapter 4: Combating salinity stress through masking effect of *miR166* on *OsMYB6*

Methodology: miRNA Sequence Extraction &

Target Prediction

The mature sequence of *miR166* was extracted using specialized plant databases, including miRBase and psRNATarget/TarDB. Target prediction involved screening for potential mRNA interactors, analysing miRNA- and target-aligned fragments, and determining modes of inhibition. Only rice-specific targets were retained, and all gene/protein names were standardized using the UniProtKB database. Literature validation ensured the accuracy and reliability of the curated dataset.

Structural Prediction of miRNA–mRNA Duplexes

Secondary structures of miR166 and its target mRNAs were predicted using the RNAfold web server. These structures were further processed with RNAComposer to generate threedimensional models of miRNA–mRNA duplexes. The resulting tertiary structures were visualized and analysed using Discovery Studio Visualizer, enabling detailed examination of their molecular interactions.

Structural Extraction of Gene Regulator Protein (AGO1)

The three-dimensional structure of the AGO1 protein (PDB ID: 4GOP, Chain A) was retrieved from the Protein Data Bank. The structure was refined using Discovery Studio Visualizer 3.5including modelling missing residues, adding hydrogen atoms, and optimizing geometry-to ensure high-quality input for docking studies.

Molecular Docking of miRNA–mRNA Duplex and AGO Protein

Protein–nucleic acid docking was performed with the pyDockDNA web server, which utilizes FTDOCK and Amber94 force field parameters to predict and score potential binding poses between AGO1 and miRNA–mRNA duplexes. The best-scoring complexes were selected for in-depth interaction analysis using Discovery Studio Visualizer and the PLIP server to annotate key non-covalent interactions.

Result

The study demonstrates that various Osa-*miR166* family members redundantly target the same mRNA via conserved mechanisms. Structural and docking analyses show that *miR166c* forms the most stable and favourable complex with AGO1, with specific nucleotide-amino acid interactions underpinning effective gene silencing in plants.